



CHALMERS

Chalmers Publication Library

Achieving the Shannon Limit with Probabilistically Shaped BICM

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

2012 IEEE International Symposium on Information Theory, ISIT 2012. Cambridge, MA, 1- 6 July 2012 (ISSN: 2157-8095)

Citation for the published paper:

Agrell, E. ; Alvarado, A. (2012) "Achieving the Shannon Limit with Probabilistically Shaped BICM". 2012 IEEE International Symposium on Information Theory, ISIT 2012. Cambridge, MA, 1- 6 July 2012 pp. 2421-2425.

<http://dx.doi.org/10.1109/ISIT.2012.6283949>

Downloaded from: <http://publications.lib.chalmers.se/publication/157199>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Achieving the Shannon Limit with Probabilistically Shaped BICM

Erik Agrell

Department of Signals and Systems
Chalmers University of Technology
SE-41296 Göteborg
Sweden
Email: agrell@chalmers.se

Alex Alvarado

Department of Engineering
University of Cambridge
Cambridge CB2 1PZ
United Kingdom
Email: alex.alvarado@ieee.org

Abstract—Probabilistic shaping for bit-interleaved coded modulation (BICM) systems at low signal-to-noise ratio (SNR) is investigated. Using known results for BICM systems with a uniform input distribution, the combinations of input alphabet, input distribution, and binary labeling that achieve the Shannon limit -1.59 dB are fully characterized. It is found that a BICM system achieves the Shannon limit at low SNR if and only if it can be represented as a zero-mean linear projection of a hypercube, which is the same condition as for uniform input distributions. Hence, probabilistic shaping offers no extra degrees of freedom to optimize the low-SNR mutual information of BICM systems, in addition to what is provided by geometrical shaping.

I. INTRODUCTION

The most important breakthrough for coded modulation (CM) in fading channels came in 1992, when Zehavi introduced the so-called bit-interleaved coded modulation (BICM) in [1], usually referred to as a pragmatic approach for CM [2], [3]. Despite not being fully understood theoretically, BICM has been rapidly adopted in commercial systems such as wireless and wired broadband access networks, 3G/4G telephony, and digital video broadcasting, making it as the de facto standard for current telecommunications systems.

Signal shaping [4], [5] refers to the use of non-equally spaced and/or non-equally likely symbols, i.e., *geometrical shaping* and *probabilistic shaping*. In the context of BICM, geometrical shaping for BICM was proposed in [6]–[8] and probabilistic shaping in [9].

For the additive white Gaussian noise (AWGN) channel, the Shannon Limit (SL) -1.59 dB represents the average bit energy to noise-ratio needed to transmit information reliably when the signal-to-noise ratio (SNR) tends to zero [10], [11]. The SL is fully determined by the first derivative of the mutual information (MI) function at zero SNR [10, Sec I] [11, eq. (1)].

When discrete alphabets are used with a BICM receiver, the SL is not always achieved [12]. This was shown in [13] to be caused by the selection of the binary labeling. The behavior of the BICM-MI at low SNR, i.e., in the wideband regime, was studied in [12]–[15]. In [16], the wideband behavior was shown to be determined by three transmission parameters:

the input alphabet \mathbb{X} , its input distribution \mathbb{P} , and its binary labeling \mathbb{L} . First-order optimal (FOO) constellations were defined as the triplet $[\mathbb{X}, \mathbb{P}, \mathbb{L}]$ that make a BICM system achieve the SL. The optimality analysis in [16], however, only considers uniform input distributions. Probabilistic shaping for BICM was considered in [9], [17]–[19]. Recently, the first-order equivalence between some uniform and nonuniform constellations for BICM was numerically studied in [20].

The main contribution of this paper is to generalize the results of [16] to nonuniform input distributions using the equivalence in [20]. A complete characterization of FOO constellations for BICM in terms of $[\mathbb{X}, \mathbb{P}, \mathbb{L}]$ is obtained. In particular, we find the geometrical and/or probabilistic shaping rules that should be applied to a constellation to make it FOO. We conclude that for BICM systems in the wideband regime, probabilistic shaping offers no extra degrees of freedom in addition to what is provided by geometrical shaping.

II. PRELIMINARIES

A. Notation Convention

Throughout this paper all MIs are given in bits. Vectors are row vectors \mathbf{x} and matrices are denoted by \mathbb{X} . The inner product between two vectors \mathbf{a} and \mathbf{b} is denoted by $\langle \mathbf{a}, \mathbf{b} \rangle$. The Euclidean norm of the vector \mathbf{a} is denoted by $\|\mathbf{a}\|$. Random variables are denoted by capital letters X and random vectors by boldface capital vectors \mathbf{X} . The probability density function (pdf) of the random vector \mathbf{Y} is denoted by $p_{\mathbf{Y}}(\mathbf{y})$ and the conditional pdf by $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$. A similar notation applies to probability mass functions (pmf) of a random variable, which we denote by $P_{\mathbf{Y}}(\mathbf{y})$ and $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$. Expectations are denoted by \mathbb{E} . The binary set by $\mathcal{B} \triangleq \{0, 1\}$. The negation of a bit b is denoted by $\bar{b} = 1 - b$.

B. System Model

We consider transmissions over a discrete-time memoryless AWGN channel. The received vector at any discrete time instant is $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, where \mathbf{X} is the channel input and \mathbf{Z} is a Gaussian noise with zero mean and variance $N_0/2$ in each dimension. The conditional transition pdf is given by

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(N_0\pi)^{N/2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{x}\|^2}{N_0}\right). \quad (1)$$

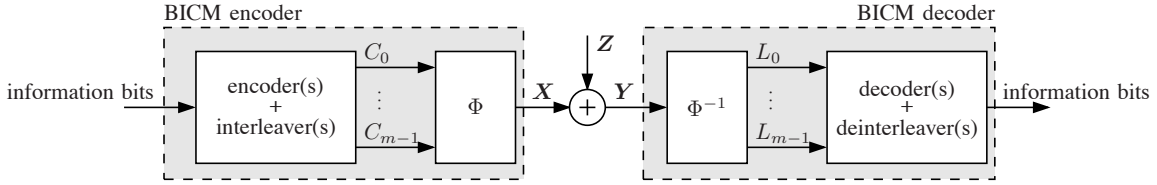


Fig. 1. A generic BICM system, consisting of a transmitter, an AWGN channel, and a receiver.

The SNR is defined as $\text{snr} \triangleq E_s/N_0 = R_c E_b/N_0$, where $E_s \triangleq \mathbb{E}_{\mathbf{X}}[\|\mathbf{X}\|^2]$ is the average symbol energy, R_c is the transmission rate in information bits per symbol, and E_b is the average energy per information bit.

We consider the generic BICM scheme in Fig. 1. In the simplest case, a binary encoder is concatenated with a memoryless modulator Φ via a bit-level interleaver. Multiple encoders and/or interleavers may be used to obtain probabilistic shaping. At the receiver, using the channel output \mathbf{Y} , the demapper Φ^{-1} computes metrics L_k for the individual coded bits C_k with $k = 0, \dots, m-1$, usually in the form of logarithmic likelihood ratios. These metrics are then passed to the deinterleaver(s) and decoder(s) to obtain an estimate of the information bits.

The mapper Φ is defined via the input alphabet $\mathbb{X} = [\mathbf{x}_0^T, \dots, \mathbf{x}_{M-1}^T]^T \in \mathbb{R}^{M \times N}$, where m bits are used to index the symbols vectors $\mathbf{x}_i \in \mathbb{R}^N$ for $i = 0, \dots, M-1$, $M = 2^m$. We associate with each symbol \mathbf{x}_i the codeword (binary labeling) $\mathbf{c}_i \triangleq [c_{i,0}, \dots, c_{i,m-1}] \in \mathbb{B}^m$ and the probability $0 \leq P_i \leq 1$, where $P_i \triangleq P_{\mathbf{X}}(\mathbf{x}_i)$. The binary labeling is denoted by $\mathbb{L} = [\mathbf{c}_0^T, \dots, \mathbf{c}_{M-1}^T]^T \in \mathbb{B}^{M \times m}$ and the input distribution by $\mathbb{P} = [P_0, \dots, P_{M-1}]^T \in [0, 1]^M$.

For any constellation $[\mathbb{X}, \mathbb{P}, \mathbb{L}]$, a set of equivalent constellations can be constructed by simultaneously permuting the rows of \mathbb{X} , \mathbb{L} , and \mathbb{P} . Specifically, it is possible to permute the rows of any labeling matrix \mathbb{L} such that any other labeling is obtained. Without loss of generality, we therefore fix the labeling to be the natural binary code (NBC) from now on, defined as follows.

Definition 1 (Natural binary code): The NBC is the binary labeling $\mathbb{N}_m \triangleq [\mathbf{n}(0)^T, \dots, \mathbf{n}(M-1)^T]^T$, where $\mathbf{n}(i) = [n_{i,0}, \dots, n_{i,m-1}] \in \mathbb{B}^m$ denotes the base-2 representation of the integer $0 \leq i \leq M-1$, with $n_{i,m-1}$ being the most significant bit.

Based on the previous discussion, from now on we use the name *constellation* to denote the pair $[\mathbb{X}, \mathbb{P}]$, where the NBC labeling is implicit. Assuming independent, but possibly nonuniformly distributed, bits C_0, \dots, C_{m-1} at the input of the modulator, the symbol probabilities are given by [16, eq. (30)] [9, eq. (8)]

$$P_i = \prod_{k=0}^{m-1} P_{C_k}(n_{i,k}) \quad (2)$$

for $i = 0, \dots, M-1$ and for some $[P_{C_0}(0), \dots, P_{C_{m-1}}(0)]$. Since $P_{C_k}(1) = 1 - P_{C_k}(0)$, the distribution \mathbb{P} is fully specified by the set of bit probabilities $[P_{C_0}(0), \dots, P_{C_{m-1}}(0)]$. An important special case is the *uniform distribution*, for which

$$P_{C_k}(0) = 1/2 \text{ for } k = 0, \dots, m-1 \text{ and } \mathbb{P} = \mathbb{U}_m \triangleq [1/M, \dots, 1/M]^T.$$

Throughout this paper, we assume that $0 < P_{C_k}(0) < 1$ for all $k = 0, \dots, m-1$, i.e., we assume that all constellation points are used with a nonzero probability. If that was not the case, the cardinality of the constellation should be reduced. Hence $P_i > 0$ for $i = 0, \dots, M-1$.

C. The Hadamard Transform

The Hadamard transform (HT), or Walsh–Hadamard transform, is a discrete, linear, orthogonal transform, whose coefficients take values in ± 1 . It is defined as follows.

Definition 2: The HT $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_0^T, \dots, \tilde{\mathbf{x}}_{M-1}^T]^T$ of a matrix (or vector) $\mathbf{X} = [\mathbf{x}_0^T, \dots, \mathbf{x}_{M-1}^T]^T$ with $M = 2^m$ rows is

$$\tilde{\mathbf{x}}_i \triangleq \frac{1}{M} \sum_{j=0}^{M-1} \mathbf{x}_j h_{i,j}, \quad i = 0, \dots, M-1 \quad (3)$$

where for all $i = 0, \dots, M-1$ and $j = 0, \dots, M-1$

$$h_{i,j} = \prod_{k=0}^{m-1} (-1)^{n_{i,k} n_{j,k}}. \quad (4)$$

Because $n_{0,k} = 0$ for $k = 0, \dots, m-1$, setting $i = 0$ in (3)–(4) shows that the first HT vector

$$\tilde{\mathbf{x}}_0 = \frac{1}{M} \sum_{j=0}^{M-1} \mathbf{x}_j \quad (5)$$

can be interpreted as the uniformly weighted mean of the alphabet. This is a property that the HT shares with, e.g., the discrete Fourier transform.

The inverse transform is identical to the forward transform, apart from a scale factor:

$$\mathbf{x}_j = \sum_{i=0}^{M-1} \tilde{\mathbf{x}}_i h_{i,j}, \quad j = 0, \dots, M-1. \quad (6)$$

III. BICM AT LOW SNR

A. Mutual Information

The *mutual information* (MI) in bits per channel use between the random vectors \mathbf{X} and \mathbf{Y} is defined as

$$I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}) \triangleq \mathbb{E} \left[\log_2 \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})}{p_{\mathbf{Y}}(\mathbf{Y})} \right], \quad (7)$$

where the expectation is taken over the joint pdf $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$, and $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ is given by (1).

The *conditional MI* is defined as the MI between \mathbf{X} and \mathbf{Y} conditioned on the value of the k th bit at the input of the modulator, i.e.,

$$I_{\mathbf{X}|C_k=u}(\mathbf{X}; \mathbf{Y}) \triangleq \mathbb{E} \left[\log_2 \frac{p_{\mathbf{Y}|\mathbf{X}, C_k}(\mathbf{Y}|\mathbf{X}, u)}{p_{\mathbf{Y}|C_k}(\mathbf{Y}|u)} \right], \quad (8)$$

where the expectation is taken over the conditional joint pdf $p_{\mathbf{X}, \mathbf{Y}|C_k}(\mathbf{x}, \mathbf{y}|u)$.

Definition 3 (BICM Mutual Information): The BICM mutual information (BICM-MI) is defined as [2], [12], [13], [21]

$$I(\text{snr}) \triangleq m I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}) - \sum_{k=0}^{m-1} \sum_{u \in \mathcal{B}} P_{C_k}(u) I_{\mathbf{X}|C_k=u}(\mathbf{X}; \mathbf{Y}). \quad (9)$$

Martinez *et al.* [21] recognized the BICM decoder in Fig. 1 as a mismatched decoder and showed that the BICM-MI in (9) corresponds to an achievable rate of such a decoder. This means that reliable transmission using a BICM system at rate R_c is possible if $R_c \leq I(\text{snr})$.

Definition 4 (Low-MI Parameters): The *low-MI parameters* of a constellation $[\mathbb{X}, \mathbb{P}]$ are defined as $[\boldsymbol{\mu}, E_s, \alpha]$, where

$$\begin{aligned} \boldsymbol{\mu} &\triangleq \mathbb{E}_{\mathbf{X}}[\mathbf{X}] \\ E_s &\triangleq \mathbb{E}_{\mathbf{X}}[\|\mathbf{X}\|^2] \\ \alpha &\triangleq \lim_{\text{snr} \rightarrow 0^+} \frac{I(\text{snr})}{\text{snr}}. \end{aligned}$$

It can be shown that for asymptotically low SNR, i.e., in the wideband regime, the average received bit energy needed for reliable transmission is lower-bounded by $E_b/N_0 \geq 1/\alpha \geq \log_e 2$, where the bound $\log_e 2$ equals the SL -1.59 dB.

B. Low-MI Parameters for Uniform Distributions

The low-MI parameters $[\boldsymbol{\mu}, E_s, \alpha]$ have been analyzed in detail for arbitrary input alphabets \mathbb{X} under the assumption of uniform probabilities [16]. Under this assumption, they can be expressed as functions of the HT of the alphabet $\tilde{\mathbb{X}}$ as shown in the following theorem.

Theorem 1: For a constellation $[\mathbb{X}, \mathbb{U}_m]$, the low-MI parameters are

$$\boldsymbol{\mu} = \frac{1}{M} \sum_{i=0}^{M-1} \mathbf{x}_i = \tilde{\mathbf{x}}_0, \quad (10)$$

$$E_s = \frac{1}{M} \sum_{i=0}^{M-1} \|\mathbf{x}_i\|^2 = \sum_{i=0}^{M-1} \|\tilde{\mathbf{x}}_i\|^2, \quad (11)$$

$$\alpha = \frac{\log_2 e}{M^2 E_s} \sum_{k=0}^{m-1} \left\| \sum_{i=0}^{M-1} (-1)^{n_{i,k}} \mathbf{x}_i \right\|^2 = \frac{\log_2 e}{E_s} \sum_{k=0}^{m-1} \|\tilde{\mathbf{x}}_{2^k}\|^2. \quad (12)$$

Proof: The expressions in (10) are obtained from Definition 4 and (5), the ones in (11) from Definition 4 and [16, eq. (16)], and the ones in (12) from [16, eq. (50)] and [16, Theorem 11]. \square

C. Low-MI Parameters for Nonuniform Distributions

Theorem 1 can be generalized to arbitrary constellations $[\mathbb{X}, \mathbb{P}]$ as follows.

Theorem 2: For a constellation $[\mathbb{X}, \mathbb{P}]$, the low-MI parameters are

$$\boldsymbol{\mu} = \sum_{i=0}^{M-1} P_i \mathbf{x}_i, \quad (13)$$

$$E_s = \sum_{i=0}^{M-1} P_i \|\mathbf{x}_i\|^2, \quad (14)$$

$$\alpha = \frac{\log_2 e}{E_s} \sum_{i=0}^{M-1} P_i \sum_{j=0}^{M-1} P_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \cdot \sum_{k=0}^{m-1} (-1)^{n_{i,k} + n_{j,k}} \frac{P_{C_k}(\tilde{n}_{i,k})}{P_{C_k}(n_{j,k})}. \quad (15)$$

Proof: Expressions (13) and (14) follow from Definition 4, while (15) can be obtained using [16, Theorem 10] plus some algebra. For a detailed proof, we refer the reader to [22, Theorem 6]. \square

Theorem 2 shows that the low-MI parameters depend on the input alphabet \mathbb{X} , the binary labeling (via $n_{i,k}$ in the expression for α), and the input distribution (via $P_{C_k}(u)$ and P_i).

IV. FOO CONSTELLATIONS FOR BICM

Having characterized the low-SNR behavior of the BICM-MI of an arbitrary constellation, the next step is to search for optimal constellations in terms of the BICM-MI at low SNR. The following definition formally defines BICM systems that achieve the SL.

Definition 5 (FOO constellation): The constellation $[\mathbb{X}, \mathbb{P}]$ is said to be first-order optimal (FOO) if a BICM system using $[\mathbb{X}, \mathbb{P}]$ achieves the SL -1.59 dB, i.e., $\alpha = \log_2 e$.

A. FOO Constellations for Nonuniform Distributions

In what follows, we derive necessary and sufficient conditions for a BICM system, with an arbitrary input alphabet and probability distribution, to achieve the SL. The proofs for the next two theorems, which are omitted due to space limitations, rely on a linear transform introduced in [20]. We refer the reader to [22] for them.

Theorem 3 ([22, Theorem 9]): The constellation $[\mathbb{X}, \mathbb{P}]$ is FOO if and only if the HT $\tilde{\mathbb{X}}$ of \mathbb{X} satisfies both the following conditions:

$$\tilde{\mathbf{x}}_0 = \sum_{k=0}^{m-1} \tilde{\mathbf{x}}_{2^k} (P_{C_k}(1) - P_{C_k}(0)) \quad (16)$$

$$\tilde{\mathbf{x}}_j = \mathbf{0}, \quad \forall j \notin \{0\} \cup \{1, 2, 4, \dots, M/2\}. \quad (17)$$

Note that only (16) depends on the input distribution, not (17). In view of [16, Theorem 12], the only difference between FOO constellations with uniform and nonuniform distributions lie in $\tilde{\mathbf{x}}_0$. The final theorem gives this statement a more intuitive interpretation.

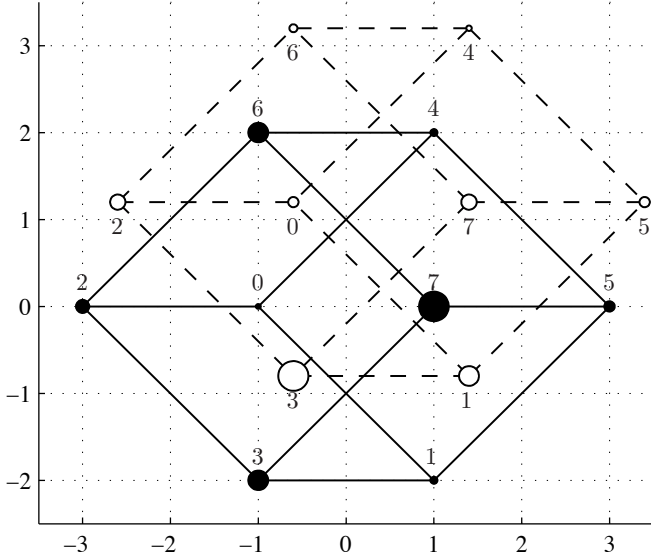


Fig. 2. Constellations $[\mathbb{X}_{\text{OTTO}}, \mathbb{P}_1]$ (filled circles) and $[\mathbb{X}'_{\text{OTTO}}, \mathbb{P}_2]$ (empty circles). Since both constellations are zero mean, they are FOO. The area of the circles is proportional to the corresponding symbol probabilities. Lines join symbols whose labels differ in one bit only.

Theorem 4 ([22, Theorem 10]): The constellation $[\mathbb{X}, \mathbb{P}]$ is FOO if and only if both the following conditions hold:

$$\boldsymbol{\mu} = \mathbf{0} \quad (18)$$

$$\tilde{\mathbf{x}}_j = \mathbf{0}, \quad \forall j \notin \{0\} \cup \{1, 2, 4, \dots, M/2\}. \quad (19)$$

Based on [16, Theorem 12], the result in Theorem 4 can be understood as follows. If a constellation with a uniform input distribution is FOO, it will still be FOO for any other input distribution $P_{C_0}(0), \dots, P_{C_{m-1}}(0)$ provided that the input alphabet is translated to be zero mean. The result in Theorem 4 also states that a constellation is FOO if and only if its input alphabet is a linear projection of a hypercube and it has zero mean. This behavior is illustrated in Example 1 and also in [16, Fig. 4].

We also note that the zero-mean condition in Theorem 4 is the same that guarantees FOO for the coded modulation MI (CM-MI) [16, Footnote 12]¹. This implies that the only difference between FOO constellations for the CM-MI and the BICM-MI lies on the extra constraint on the input alphabet to be a linear projection of a hypercube.

B. Numerical Examples

In this section we give numerical examples to illustrate the analytical results presented in this paper.

Example 1: Consider the so-called OTTO (one-three-three-one) alphabet shown in [16, Fig. 4 (a)], which corresponds to a projected hypercube. This alphabet is

$$\mathbb{X}_{\text{OTTO}} = \begin{bmatrix} -1 & 1 & -3 & -1 & 1 & 3 & -1 & 1 \\ 0 & -2 & 0 & -2 & 2 & 0 & 2 & 0 \end{bmatrix}^T. \quad (20)$$

¹The parameter α for the CM-MI is [16, Theorem 7] $\alpha = \log_2 e(1 - \|\boldsymbol{\mu}\|^2/E_s)$.

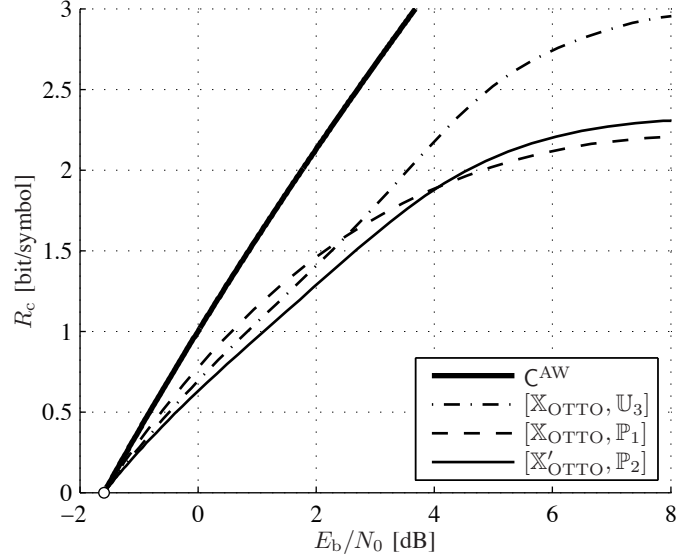


Fig. 3. BICM-MI for the two FOO constellations in Fig. 2. The BICM-MI for the constellation $[\mathbb{X}_{\text{OTTO}}, \mathbb{U}_3]$ in [16, Fig. 4 (a)] and the AWG capacity $C_{\text{AW}} = \log_2(1 + E_s/N_0)$ are shown for comparison. The SL is shown with a white circle.

The constellation $[\mathbb{X}_{\text{OTTO}}, \mathbb{U}_3]$ was shown to be FOO in [16, Example 4].

In this example, we are interested in the first-order behavior of the constellation $[\mathbb{X}_{\text{OTTO}}, \mathbb{P}]$ for different \mathbb{P} . In view of Theorem 4, this constellation will be FOO if it has zero mean. Using (20) and (2) in (13), we find (after some algebra) that

$$\boldsymbol{\mu} = \begin{bmatrix} 1 + 2(P_{C_1}(0) - P_{C_0}(0) - P_{C_2}(0)) \\ 2(P_{C_0}(0) - P_{C_2}(0)) \end{bmatrix}^T. \quad (21)$$

For certain distributions, the mean (21) is zero. Specifically, $\boldsymbol{\mu} = \mathbf{0}$ if and only if

$$P_{C_0}(0) = P_{C_2}(0) = P_{C_1}(0)/2 + 1/4. \quad (22)$$

Clearly, the uniform case ($P_{C_0}(0) = P_{C_1}(0) = P_{C_2}(0) = 1/2$) analyzed in [16] fulfills (22). When any other set of bit probabilities fulfilling (22) is used, the resulting constellation will be FOO. This is the case for instance if $[P_{C_0}(0), P_{C_1}(0), P_{C_2}(0)] = [0.3, 0.1, 0.3]$ which induces via (2) an input distribution denoted by \mathbb{P}_1 . The obtained constellation $[\mathbb{X}_{\text{OTTO}}, \mathbb{P}_1]$ is illustrated in Fig. 2 with filled circles, where the area of the circles is proportional to the corresponding symbol probabilities.

Input distributions do not generally fulfill (22). For example, the set of probabilities $[P_{C_0}(0), P_{C_1}(0), P_{C_2}(0)] = [0.2, 0.3, 0.8]$ gives an input distribution \mathbb{P}_2 for which the mean (21) is $\boldsymbol{\mu} = [-0.4, -1.2]$. We define a new alphabet $\mathbb{X}'_{\text{OTTO}}$ by subtracting $\boldsymbol{\mu}$ from each element in \mathbb{X}_{OTTO} . The translated constellation $[\mathbb{X}'_{\text{OTTO}}, \mathbb{P}_2]$, shown in Fig. 2, is a zero-mean projected hypercube and thus FOO according to Theorem 4.

In Fig. 3, we show the BICM-MI for the zero-mean constellations $[\mathbb{X}_{\text{OTTO}}, \mathbb{P}_1]$ and $[\mathbb{X}'_{\text{OTTO}}, \mathbb{P}_2]$ in Fig. 2 as well as for the constellation $[\mathbb{X}_{\text{OTTO}}, \mathbb{U}_3]$ in [16, Fig. 4 (a)]. As expected, all the MI curves converge at the SL for low SNR.

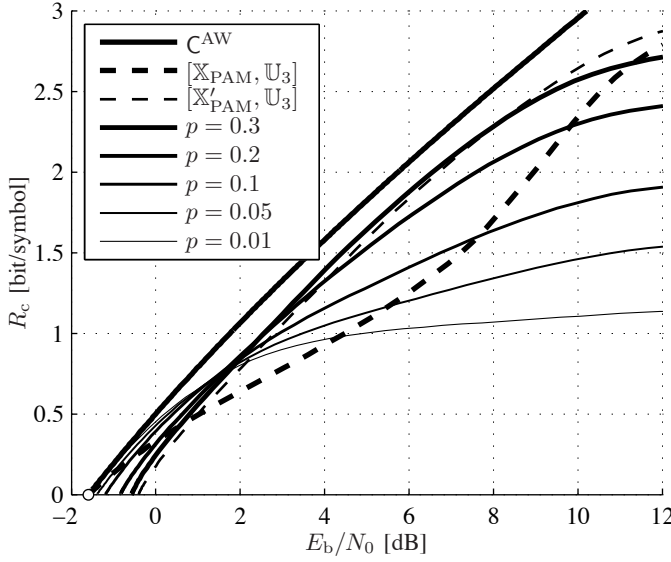


Fig. 4. The BICM-MI for a Gray-labeled 8-PAM alphabet \mathbb{X}'_{PAM} with uniform input distribution and bit probabilities $[P_{C_0}(0), P_{C_1}(0), P_{C_2}(0)] = [0.5, p, p]$ for different values of p . The constellations approach a binary FOO constellation as $p \rightarrow 0$. The NBC-labeled 8-PAM constellation $[\mathbb{X}_{\text{PAM}}, \mathbb{U}_3]$, although FOO, is considerably weaker than \mathbb{X}'_{PAM} for a wide range of SNRs.

The exemplified method holds in full generality: Any alphabet that is FOO with a uniform input distribution is FOO also with an arbitrary nonuniform distribution, if it is translated to zero mean. Furthermore, *all* nonuniform FOO constellations can be constructed in this manner.

Example 2: M -ary pulse amplitude modulation (PAM) alphabets have been shown to be FOO if the NBC is used with a uniform input distribution, i.e., the constellation $[\mathbb{X}_{\text{PAM}}, \mathbb{U}_m]$ with $\mathbb{X}_{\text{PAM}} = [-(M-1), -(M-3), \dots, M-1]$ is FOO [13] [16, Theorem 14]. In this example, we study the first-order behavior of the Gray-labeled 8-PAM alphabet $\mathbb{X}'_{\text{PAM}} = [-7, 7, -1, 1, -5, 5, -3, 3]$. The constellation $[\mathbb{X}'_{\text{PAM}}, \mathbb{P}]$ is known not to be FOO for \mathbb{U}_3 [12, Theorem 3].

The BICM-MI of \mathbb{X}'_{PAM} is shown in Fig. 4 for the set of bit probabilities $[P_{C_0}(0), P_{C_1}(0), P_{C_2}(0)] = [0.5, p, p]$ for different values of p . For $p = 0.5$, the uniform distribution is obtained. As p decreases, the Gray-coded constellation approaches a zero-mean binary alphabet, which is FOO [9, Fig. 2] [16, Fig. 3(b)]. The results in Fig. 4 show the tradeoff between the low- and high-SNR regimes: The SL can be approached by decreasing p , but this causes a decrease in MI in the high-SNR regime. Alternatively, the SL can be attained by switching from the Gray code to the NBC, but this also comes with a heavy penalty at higher SNRs.

V. CONCLUSIONS

We have characterized BICM systems that attain the SL -1.59 dB at asymptotically low SNR, i.e., FOO constellations. Somewhat disappointingly, the set of probabilistically shaped FOO constellation is no larger than the set of FOO constellations with uniform distributions, disregarding translations of

the whole input alphabet. Both sets can be fully characterized as the set of linear projections of a hypercube, translated to have zero mean for the considered input distribution. It is impossible to make non-FOO constellations for BICM FOO by probabilistic shaping, except in degenerate cases (by setting some probabilities equal to zero).

REFERENCES

- [1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.
- [2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [3] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.
- [4] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 726–740, July 1990.
- [5] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, 2002.
- [6] D. Sommer and G. P. Fettweis, "Signal shaping by non-uniform QAM for AWGN channels and applications using turbo coding," in *International ITG Conference on Source and Channel Coding (SCC)*, Munich, Germany, Jan. 2000.
- [7] S. Y. Le Goff, "Signal constellations for bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 307–313, Jan. 2003.
- [8] M. Barsoum, C. Jones, and M. Fitz, "Constellation design via capacity maximization," in *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007.
- [9] A. Guillén i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, Aug.–Sep. 2010.
- [10] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
- [11] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.
- [12] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Bit-interleaved coded modulation in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5447–5455, Dec. 2008.
- [13] C. Stierstorfer and R. F. H. Fischer, "Asymptotically optimal mappings for BICM with M -PAM and M^2 -QAM," *IET Electronics Letters*, vol. 45, no. 3, pp. 173–174, Jan. 2009.
- [14] A. Alvarado, E. Agrell, A. Guillén i Fàbregas, and A. Martinez, "Corrections to 'Bit-interleaved coded modulation in the wideband regime'," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, p. 6513, Dec. 2010.
- [15] C. Stierstorfer and R. F. H. Fischer, "Mappings for BICM in UWB scenarios," in *International ITG Conference on Source and Channel Coding (SCC)*, Ulm, Germany, Jan. 2008.
- [16] E. Agrell and A. Alvarado, "Optimal alphabets and binary labelings for BICM at low SNR," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6650–6672, Oct. 2011.
- [17] D. Raphaeli and A. Gurevitz, "Constellation shaping for pragmatic turbo-coded modulation with high spectral efficiency," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 341–345, Mar. 2004.
- [18] S. Y. Le Goff, B. S. Sharif, and S. A. Jima, "Bit-interleaved turbo-coded modulation using shaping coding," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 246–248, Mar. 2005.
- [19] S. Y. Le Goff, B. K. Khoo, C. C. Tsimenidis, and B. S. Sharif, "Constellation shaping for bandwidth-efficient turbo-coded modulation with iterative receiver," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2223–2233, June 2007.
- [20] E. Agrell and A. Alvarado, "First-order asymptotics of the BICM mutual information: Uniform vs. nonuniform distributions," in *Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, Feb. 2012.
- [21] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.
- [22] E. Agrell and A. Alvarado, "Signal shaping for BICM at low SNR," [Online]. Available: <http://arxiv.org>, Feb. 2012.